Health Psychology

Area-Level Racial Prejudice and Health: A Systematic Review

Eli K. Michaels[1], Christine Board[1], Mahasin S. Mujahid[1], Corinne A. Riddell[1,2], David H. Chae[3], Rucker C. Johnson[4], and Amani M. Allen[1,5]

[1] Division of Epidemiology, School of Public Health, University of California, Berkeley

[2] Division of Biostatistics, School of Public Health, University of California, Berkeley

[3] Department of Global Community Health & Behavioral Sciences, Tulane School of Public Health and Tropical Medicine

[4] Goldman School of Public Policy, University of California, Berkeley

[5] Division of Community Health Sciences, School of Public Health, University of California, Berkeley

Author Note

Eli K. Michaels ORCID: https://orcid.org/0000-0001-9209-2560

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Eli K. Michaels, MPH, School of Public Health, University of California, Berkeley, 2121 Berkeley Way, Room 5302, Berkeley, CA 94720-7360, Email: elikmichaels@berkeley.edu

Appendix A: Detailed Methodology

Search Strategy

We searched the following electronic databases with the goal of gathering studies from across academic disciplines: (1) PubMed, (2) SCOPUS, (3) PsycInfo, and (4) Sociological Abstracts.

Search terms were developed iteratively based on a preliminary review of the literature, research team expertise, content knowledge, and consultation with a public health research librarian. First, we developed preliminary search terms based on the titles and abstracts of known twelve papers examining the association between area-level racial prejudice and health outcomes (Chae et al., 2015; Chae et al., 2018; Hehman et al., 2018; Huang et al., 2020; Kennedy et al., 1997; Lee et al., 2015; Leitner et al., 2016a, 2016b; McKetta et al., 2017; Morey et al., 2018; Nguyen et al., 2018; Orchard & Price, 2017). Next, we added search terms identified by the research team and those recommended by the research librarian. We then tested preliminary search strings in multiple databases to gauge the breadth and depth of results returned. We iteratively modified search terms, string combinations, and databases to ensure all twelve known papers were identified. Once our search strategy identified all twelve known papers, we performed the formal search with no further modifications. The final set of strings were:

STRING 1: “racism” OR “stigma” OR “racial prejudice” OR “racial bias” OR “racial biases” OR “implicit racial bias” OR “explicit racial bias” OR “racial attitudes” OR “racist attitudes” OR “racial beliefs” OR “racist beliefs” OR “racial sentiment” OR “racist sentiment” OR “N-Word” OR “racial animus”

STRING 2: "project implicit" OR "general social survey" OR “Twitter” OR “Google”

STRING 3: “community-level” OR “communities” OR “county-level” OR “state-level” OR "area-level" OR “neighborhood-level” OR “regional” OR “collective”

STRING 4: “area-racism” OR “collective disrespect” OR “bias of crowds”
QUERY 1: string 1 AND string 2
QUERY 2: string 1 AND string 3
QUERY 3: string 2 AND string 3
QUERY 4: string 4

We performed our database search on April 5, 2020. One investigator entered search strings into databases 1 and 2, and another investigator entered search strings into databases 3 and 4. Our search yielded a total of 20,616 records, which were uploaded to Covidence systematic review software (Innovation, 2016). Two articles, published in July and September of 2020, were identified after the formal literature pull but before data extraction was complete (Hswen, 2020; Nguyen et al., 2020). We included these papers for consideration in the review to maximize the amount of information gained from this emerging area of research. After removal of duplicates, 14,632 records proceeded to title and abstract screening.

Two reviewers independently performed all screening based on inclusion and exclusion criteria. Results of the screening were compared, and disagreements were resolved via consult from a third investigator. Inclusion criteria included: (a) peer-reviewed journal article; (b) quantitative empirical study; (c) US-based; (d) English language; (e) study exposure is an indicator of bias, prejudice, animus, attitudes, sentiment, or beliefs toward or about a particular racial, ethnic, or immigrant group(s) that is measured at the individual-level and aggregated to the area-level; (f) study exposure is assessed using data from (1) surveys, questionnaires, or assessment tools, (2) social media, or (3) Google searches; and (g) study outcome is a mental or physical health outcome or health behavior.

Our title and abstract screening excluded 14,600 records, leaving 32 articles for full-text review, of which 14 met inclusion criteria. Figure 1 shows the results of these exclusions.

Data Extraction

Once the final set of included papers was identified, full-text PDFs were uploaded into MaxQDA (Software, 2019) for data extraction.

We extracted standard data in accordance with PRISMA guidelines.(Moher et al., 2009) We also extracted data for our specific research questions. First, we were interested in conceptualization and framing—how were researchers thinking about area-level racial prejudice in relation to existing conceptual models for racism and health? We documented the terminology and theory used to describe the exposure, presence and content of any conceptual models, and proposed pathways to health. Second, we extracted data on empirically tested mediation and moderation of the association between area-level racial prejudice and health outcomes. In particular, we were interested in whether any association between area-level racial prejudice and health outcomes was differential by racial identity. Finally, we extracted data on key measurement and other methodological considerations.

Data Extraction Codebook

1 Background/Framing

1.1 Motivation

How are the authors motivating their approach to aggregating racial bias (e.g., to measure structural/cultural racism, to avoid self-report, some other reason, no rationale provided?)

1.2 Terminology

Terminology used to describe the exposure

1.3 Theory

Theory used? If so, which theory or theories?

1.4 Conceptual model

1.5 Pathway to health

Proposed pathway to health?

2 Study population

2.1 Exposure geography

Number of geographic units in exposure population (e.g., 208 DMAs)

2.2 Outcome pop

Number and demographic breakdown (e.g., age, sex, other) of participants in outcome assessment (e.g., outcomes on 40,000 NHW and NHB BRFSS respondents)

3 Study design

3.1 Follow-up

One time-point (cross-sectional)

Multiple time points (longitudinal

Time-to-event (survival)

3.2 Level of analysis

Ecologic - exposure and outcome measured at area-level

Multilevel - exposure at area-level (accounts for clustering), outcome at individual-level

Individual - exposure and outcome at individual-level (does not account for clustering)

3.3 Study period

Time period of data (exposure, outcome, and covariates)

4 Study setting

4.1 Study area

E.g., California, US, global

4.2 Exposure scale

At what geographic scale was the exposure measured?

4.3 Outcome scale

At what geographic scale was the outcome measured?

4.4 Covariate scale

At what geographic scale were covariates measured

5 Exposure

5.1 Exposure(s)

5.2 Data source

Project Implicit

Google

General Social Survey

Twitter

Other

5.3 Number aggregated

Number of individual observations aggregated (e.g., n=1 million IAT responses were aggregated to the county-level) (if reported)

5.4 Specification

Implicit or explicit racial bias data

Restrictions

Weighting

Google search terms queried

Continuous or binary

Coding, cutpoints, etc.

Any information on validity (either based on prior literature, or tested in the study)

6 Outcomes

6.1 Outcome(s)

What was the primary study outcome?

6.2 Data source

What was the data source for the study outcome?

6.3 Assessment

E.g., self-report, biomarker, administrative records

6.4 Specification

E.g., Continuous, binary coding/cutpoints used; other details

7 Confounder adjustment

7.1 How identified?

How were confounders identified (e.g., literature review, DAG, data-driven approaches)?

7.2 Area-level

What area-level confounders were identified and how were they measured?

7.3 Individual-level

What individual-level confounders were identified and how were they measured?

7.4 How addressed?

How was confounding addressed (e.g., multivariable regression, propensity score matching, econometric models?)?

8 Findings

8.1 Statistical model

Statistical model used and any modeling notes (e.g., robust SEs, sensitivity analyses performed, etc)

8.2 MOA

Measure of association and 95% confidence interval for main results, subgroup effects, and any sensitivity analyses

8.3 Findings

Tag to highlight summary of findings

9 Mechanisms

9.1 Area-level

Area-level mediation or effect measure modification (formal interaction or stratified results)

9.2 Individual-level

Individual-level mediation or effect measure modification (formal interaction or stratified results)

9.3 Differential?

Association differential or non-differential by racial identity (assessed via formal interaction or race-stratified results)?

10 Limitations

10.1 Investigator

Limitations identified by the investigator

10.2 Research team

Limitations identified by the research team

11 Notes

11.1 Implications

Implications for future research

11.2 Other refs

Any other references to include in the review that we missed in our literature pull

11.3 Notable

Anything else you find notable or want to come back to; memorable quotes

Appendix B. Detailed Study Information

Characteristics of Included Studies

Table 1 displays the characteristics of studies included in the systematic review. Ten studies were cross-sectional (i.e., exposure and outcome examined at one time-point only, even if the exposure preceded the outcome) (Chae et al., 2015; Chae et al., 2018; Hehman et al., 2018; Huang et al., 2020; Kennedy et al., 1997; Leitner et al., 2016a, 2016b; Nguyen et al., 2020; Nguyen et al., 2018; Orchard & Price, 2017), three studies examined outcomes prospectively using survival methods (i.e., exposure precedes outcome and there are multiple outcome assessments on each study participant) (Lee et al., 2015; McKetta et al., 2017; Morey et al., 2018), and one employed a time-series analysis (i.e., changes in group-level rates over time) (Hswen, 2020). All studies measured the exposure at the area-level, but the geographic scale ranged from the county (n=3) (Leitner et al., 2016a, 2016b; Orchard & Price, 2017) to the national level (n=1) (Hswen, 2020), with the largest number of studies (n=5) examining racial prejudice at the state-level (Huang et al., 2020; Kennedy et al., 1997; McKetta et al., 2017; Nguyen et al., 2020; Nguyen et al., 2018). Seven studies were multilevel, examining health outcomes at the individual-level using analytic methods that account for clustered data (Chae et al., 2018; Lee et al., 2015; McKetta et al., 2017; Morey et al., 2018; Nguyen et al., 2020; Nguyen et al., 2018; Orchard & Price, 2017), whereas one study did not account for clustering (i.e., individual-level study) (Huang et al., 2020). The remaining six studies were ecologic with the geographic area as the unit of analysis (e.g., rates as study outcome) (Chae et al., 2015; Hehman et al., 2018; Hswen, 2020; Kennedy et al., 1997; Leitner et al., 2016a, 2016b).

Area-level racial prejudice was examined using one of four data sources: the GSS (n=3) (Kennedy et al., 1997; Lee et al., 2015; Morey et al., 2018), Project Implicit (n=4) (Hehman et al., 2018; Leitner et al., 2016a, 2016b; Orchard & Price, 2017), Google Trends (n=3) (Chae et al., 2015; Chae et al., 2018; McKetta et al., 2017), and Twitter (n=4) (Hswen, 2020; Huang et al., 2020; Nguyen et al., 2020; Nguyen et al., 2018). These data sources are described in detail in Section 3.2. Several studies specifically examined the racial bias of White (n=3) (Hehman et al., 2018; Leitner et al., 2016a, 2016b) and/or Black (n=2) (Hehman et al., 2018; Leitner et al., 2016b) respondents. Four studies did not disaggregate the exposure by respondent race

(Kennedy et al., 1997; Lee et al., 2015; Morey et al., 2018; Orchard & Price, 2017), and the remaining seven were unable to discern this information given the data available (i.e., Google or Twitter) (Chae et al., 2015; Chae et al., 2018; Hswen, 2020; Huang et al., 2020; Nguyen et al., 2020; Nguyen et al., 2018).

Studies explored a variety of health outcomes, including birth outcomes (n=4) (Chae et al., 2018; Nguyen et al., 2020; Nguyen et al., 2018; Orchard & Price, 2017), all-cause mortality (n=4) (Chae et al., 2015; Kennedy et al., 1997; Lee et al., 2015; Morey et al., 2018), cause-specific mortality (n=4) (Chae et al., 2015; Hehman et al., 2018; Leitner et al., 2016a, 2016b), cardiovascular disease (CVD) and related risk factors (n=2) (Huang et al., 2020; Leitner et al., 2016a), mental health outcomes (n=1) (Hswen, 2020), and self-rated health (n=1) (McKetta et al., 2017). Authors explored health outcomes of multiple racial/ethnic groups in relation to area-level racial prejudice. Six studies examined health outcomes of Black and White persons (Kennedy et al., 1997; Lee et al., 2015; Leitner et al., 2016a, 2016b; McKetta et al., 2017; Orchard & Price, 2017), while four studies examined health outcomes of multiple (>2) racial/ethnic groups (Huang et al., 2020; Morey et al., 2018; Nguyen et al., 2020; Nguyen et al., 2018). Three studies examined the health outcomes of Black persons only (Chae et al., 2015; Chae et al., 2018; Hehman et al., 2018), and one of Hispanics only (Hswen, 2020).

Ten of the fourteen studies explored whether associations between area-level racial prejudice and health was differential by racial/ethnic group, either by comparing race/ethnicity-specific rates (Kennedy et al., 1997; Nguyen et al., 2020; Nguyen et al., 2018) or by formally testing for multiplicative statistical interaction between area-level racial prejudice and race/ethnicity (Huang et al., 2020; Lee et al., 2015; Leitner et al., 2016a, 2016b; McKetta et al., 2017; Morey et al., 2018; Orchard & Price, 2017). The remaining four studies were restricted to one racial/ethnic group (Chae et al., 2015; Chae et al., 2018; Hehman et al., 2018; Hswen, 2020). Two studies explored mediators on the pathway from area-level racial prejudice to health outcomes (Lee et al., 2015; Leitner et al., 2016a).

Table B1. *Study Overviews*

Source	Study design		Study sample	Study setting				
First author, year	**Level of analysis**	**Follow-up**	**Study sample (descriptives if provided)**	**Study area**	**Years of data for Exposure, Outcome, Covariates**	**Exposure scale**	**Outcome scale**	**Covariate scale**
Kennedy et al., 1995	Ecologic	Cross-sectional	N/A (rates)	39 US states (not specified which states)	E: 1986-1990 O: 1990 C: 1990	Individual, aggregated to state	State	State
Lee et al., 2015	Multilevel	Prospective (discrete-time event history)	n=10,950 Black and White GSS respondents across 100 PSUs; Mage=45 years 55% female, 85.7% White, 14.3% = Black	US	E: 1993-2002 O: 1993-2008 C: 1990-2002	Individual, aggregated to PSU	Individual	Individual and PSU
Morey et al., 2018	Multilevel	Prospective (survival)	n=13,242 immigrant GSS respondents across 123 PSUs (Mage=43.5, 53% female, 79% White, 14% Black, 8% Other Race)	US	E: 1993-2010 O: 1993-2014 C: 1993-2014	Individual, aggregated to PSU	Individual	Individual and PSU
Leitner et al., 2016a	Ecologic	Cross-sectional	Study 1: n=199,159 Black and White BRFSS respondents (11.8% Black, 88.9% White) but outcomes were modeled as rates Study 2: NA (rates)	US	E: 2003-2013 O1: 2012 O2: 2003-2013 C: 2000, 2005-2013	Test, aggregated to county	Study 1: County Study 2: County	Study 1: County Study 2: County
Leitner et al., 2016b	Ecologic	Cross-sectional	N/A (rates) Black death rate per 100,000: M = 352.595, SD = 84.806; White death rate per 100,000: M = 270.477, SD = 54.2	US	E: 2003-2013 O: 2003-2013 C: 2005-2013	Test, aggregated to county	County	County
Orchard & Price, 2017	Multilevel	Cross-sectional	n=31,464,451 births (White Mage = 27.78, SD = 6.04, 15% finished college; Black Mage = 25.84, SD = 6.22, 8% finished college)	US	E: 2002-2012 O: 2002-2012 C: 2002-2013	Test, aggregated to county	Individual	Individual and county
Hehman et al., 2017	Ecologic	Cross-sectional	n=875 individuals confirmed as killed by police officers in the United States (Mage = 37.3 years, SD = 13.3; 4% female)	US	E: 2003- 2013 O: 1/1/15-9/30/15 C: 2010-2013	Test, aggregated to CBSA	CBSA	CBSA

Chae et al., 2015	Ecologic	Cross-sectional	23.1 million person-years across 196 DMAs (49.3% aged 45+; 52.81% female) but outcomes were modeled as rates	US (except AK)	E: 2004–2007 O: 2004–2009 C: 2000, 2004-2009	Search, aggregated to DMA	DMA	DMA
Chae et al., 2018	Multilevel	Cross-sectional	n=2,332,216 births to Black women across 196 DMAs (maternal age: 6.3% <18, 83.6% 18-34, 10.1% 35+)	US (except AK)	E: 2004-2007 O: 2005-2008 C: 2005-2010	Search, aggregated DMA	Individual	Individual and DMA
McKetta et al., 2018	Multilevel	Prospective (survival)	N=16,580 Black and White PSID respondents (66.1% White, 33.9% Black)	US (except AK)	E: 2004-2007 O: 1990-2009 C: 1990	DMA, aggregated to state	Individual	Individual and state
Nguyen et al., 2018	Multilevel	Cross-sectional	n=3,988,733 births (birthing persons – 53% were White, non-Hispanic, and 77% were U.S. born)	Contiguous US + DC	E: March 2015–April 2016 O: 2015 C: 2015	Tweet, aggregated to state	Individual	Individual and state
Huang et al., 2020	Individual	Cross-sectional	n=450,016 participants (range: n=433,434 to n=433,680 across outcomes)	Contiguous US + DC	E: 2015-2018 O: 2017 C: 2017	Tweet, aggregated to state	Individual	Individual
Nguyen et al., 2020	Multilevel	Serial cross-sectional	N=9,988,030 for gestational age, n=9,985,402 for birth weight (birthing persons – Mage= 29 years, 59.74% married, 85.99% completed at least high school)	Contiguous US + DC	E: June 2015-December 2017 O: 2015-2017 C: 2013-2017	Tweet, aggregated to state	Individual	Individual and state
Hswen et al., 2020	Ecologic	Time-series	n=8,314 Hispanic Gallup respondents (no descriptives provided)	US	E: 8/29/2016-1/16/2017 O: 8/29/2016-1/16/2017	Tweet, aggregated to US	Individual, aggregated to US	NA

Table B2. *Study measures*

Source	Exposure				Outcome			
First author, year	**Exposure**	**Data source**	**# aggregated**	**Operationalization/specification**	**Outcome**	**Data source**	**Assessment**	**Operationalization/specification**
Kennedy et al., 1995	Collective disrespect	GSS	n=7,679	Individual-level questions: "On average, blacks have worse jobs, income, and housing. Do you think the differences are due to (a) discrimination, (b) less in-born ability to learn, (c) lack of chance for education that it takes to rise out of poverty, (d) less motivation or willpower to pull themselves out of poverty?* * Each item was dichotomized separately Aggregate measure: state-level % of respondents who answered in the affirmative to each item Weighting: Post-stratification weights based on age, race, educational attainment Specification: continuous	Age-adjusted all-cause Black and White mortality rates	NCHS death records	Administrative (death) records	Measure: Directly age-standardized to the US population of Black and White persons, and expressed as the number of deaths per 100,000 persons. Specification: rate per 100,000
Lee et al., 2015	Community-level racial prejudice	GSS-NDI	n=13,355 (14,513 GSS respondents-1,158 with missing racial prejudice data)	Individual-level questions: 1. "On the average, negroes/blacks/African-Americans have worse jobs, income, and housing than white people. Do you think these differences are caused by the fact that most negroes/blacks/African-Americans have less in-born ability to learn?" 2. "Do you think these differences are because most negroes/blacks/African-Americans just don't have the motivation or willpower to pull themselves up out of poverty?" 3. "Do blacks tend to be unintelligent or tend to be intelligent?" (and "Do whites tend to be unintelligent or tend to be intelligent?") 4. "Do blacks tend to be hard working or lazy?" (and "Do whites tend to be hard working or lazy?") 5. "Do you think there should be laws against marriages between Negroes/Blacks/African-Americans and whites?"* * Each item was dichotomized, then averaged across items	All-cause mortality (survival)	GSS-NDI	Administrative (death) records	Measure: 0=alive in 2008, 1=died by 2008 Specification: binary (survival)

				Aggregate measure: PSU-level average scores Specification: Standardized & centered (continuous)				
Morey et al., 2018	Community-level anti-immigrant prejudice	GSS-NDI	n=2,427	Individual-level questions: 1. "Do you think the number of immigrants to America nowadays should be increased a lot, increased a little, remain the same as it is, reduced a little, or reduced a lot?" 2. Respondents were asked how much they agreed or disagreed with the following four statements: (1) "America should take stronger measures to exclude illegal immigrants," (2) "Immigrants take jobs away from people who were born in America," (3) "Immigrants increase crime rates," and (4) "Immigrants are generally good for America's economy." Responses were coded on a five-point Likert scale ranging from "agree strongly" to "disagree strongly."* * Each item was dichotomized, then summed across items Aggregate measure: PSU-level average scores Specification: Continuous and dichotomous (+/-1SD from the mean)	All-cause mortality (survival)	GSS-NDI	Administrative (death) records	Measure: 0=alive in 2014 1=died by 2014 + censored amount of time at risk over the study period Specification: binary (survival) and continuous (time-to-event)
Leitner et al., 2016a	White county-level racial bias	Project Implicit Race IAT	n=1,391,632 White IAT responses	Individual implicit measure: keyboard association test with D-score Individual explicit measure: temperature difference Aggregate measure: county-level average implicit and explicit scores of White IAT respondents Weighting: Post- stratification weights based on age Specification: Continuous and dichotomous (+/-1SD from the mean)	Study 1: Black and White circulatory-disease risk (% without access to health care*, % with circulatory disease) Study 2: Black and White age-adjusted circulatory disease mortality rates * % without access to health care does not	Study 1: BRFSS Study 2: NCHS death records	Study 1: self-report (telephone interview) Study 2: administrative (death) records	Study 1: circulatory disease risk Circulatory disease diagnosis question "Has a doctor, nurse, or other health professional ever told you that you had a heart attack, also called a myocardial infarction?" or "…angina or coronary heart disease?" Coding: 0="no," 1="yes" Aggregation: averaged at the county level to calculate county % without healthcare access and % with either diagnosis. Specification: prevalence (continuous), examined separately and as a B-W difference

					meet inclusion criteria for the review.			Study 2: circulatory disease mortality Measure: Black and White deaths from circulatory diseases (e.g., heart disease; Internal Statistical Classification of Diseases and Related Health Problems codes I00–I99). Age adjusted based on 2000 standard population (for each racial group) Specification: rates per 100,000, examined separately and as B-W difference
Leitner et al., 2016b	Black and White county-level ingroup bias	Project Implicit Race IAT	n=250,665 Black IAT responses, n=1,391,632 White IAT responses	Ingroup bias: White respondents' pro-White/anti-Black bias and Black respondents' pro-Black/anti-White bias (i.e., ingroup favoritism) Individual implicit measure: keyboard association test with D-score, scaled for ingroup Individual explicit measure: temperature difference, scaled for ingroup Aggregate measure: county-level average implicit and explicit scores Weighting: Post- stratification weights based on age Specification: Continuous and dichotomous (+/- 1SD from the mean)	Age-adjusted Black and White circulatory disease mortality rates	NCHS death records	Administrative (death) records	Measure: Black and White deaths from circulatory diseases (e.g., heart disease; Internal Statistical Classification of Diseases and Related Health Problems codes I00–I99). Age adjusted based on 2000 standard population (for each racial group) Specification: rate per 100,000
Orchard & Price, 2017	Community-level racial prejudice	Project Implicit Race IAT	n=1.8 million IAT responses aged 18+ (mean age=28, and 59% women)	Individual implicit measure: keyboard association test with D-score Individual explicit measure: preference measure Aggregate measure: county-level average implicit and explicit scores Weighting: Post-stratification weights based on age + gender [+] Specification: Standardized (continuous) and dichotomous (+/- 1SD from the mean)	Black and White rates of adverse birth outcomes	NCHS birth records	Administrative (birth) records	Measure: Binary PTB: gestational age < 37 weeks Binary LBW: < 2500 g Specification: Black and White rates per 1,000 births (B-W difference assessed with interaction term)

Hehman et al., 2017	Regional racial biases of residents (Black-White bias and weapons stereotype)	Project Implicit Race IAT and Weapons IAT	n=1,860,818 Black and White Race IAT responses, n=295,235 Black and White Weapons IAT responses	Individual implicit measure: keyboard association test with D-score (race IAT and weapons IAT) Individual explicit measure: temperature difference (race IAT) Aggregate measure: CBSA-level average implicit and explicit scores, separately for Black and White respondents Weighting: No Specification: Untransformed (continuous)	Disproportionate lethal force against Black and White people relative to their population shares	The Guardian police killing database	Traditional reporting with police reports Fact-checked witness statements; monitoring of regional news; other open-sourced police fatality databases	Measure: % of Black people living in each CBSA was subtracted from the % of Black people killed in each CBSA relative to the total amount of individuals killed by police officers. Higher score on this variable reflected greater usage of lethal force with Black people than would be expected based on the CBSA population (i.e., disproportionate lethal force). An identical score was calculated for NH White people Specification: continuous
Chae et al., 2015	Area racism	Google Trends data compiled by SSD (2014)	NA	Aggregate measure: DMA-level proportion of total Google searches containing the "n-word." (singular or plural, ending in "-er(s)" but not "-a(s)") Specification: Standardized (continuous)	Age-adjusted Black all-cause and cause-specific (heart disease, cancer, stroke, and diabetes) mortality rates	NCHS death records	Administrative (death) records	Measure: Black mortality rates weighted using the US 2000 standard population were calculated for all-cause mortality and the four leading specific causes of death among Black people identified using International Classification of Disease, Version 10 codes: heart disease (I00-I09, I11, I13, I20-I51); cancer (C00-C97); stroke (I60-I69); and diabetes (E11-E14) Specification: rate per 100,000 person-years
Chae et al., 2018	Area racism	Google Trends data compiled by SSD (2014)	NA	Aggregate measure: DMA-level proportion of total Google searches containing the "n-word." (singular or plural, ending in "-er(s)" but not "-a(s)") Specification: Standardized (continuous)	PTB and LBW among NH Black women	NCHS birth records	Administrative (birth) records	Measure: PTB: gestational age < 37 weeks LBW: < 2500 g Specification: Binary
McKetta et al., 2018	State-level racial animus	Google Trends data compiled by SSD (2014)	NA	Aggregate measure: State-level proportion of total Google searches containing the "n-word." (singular or plural, ending in "-er(s)" but not "-a(s)") (DMAs aggregated to state-level) Specification: Quartiles	Black and White SRH and Black-White differences in SRH (also movement across states)	PSID	Self-report (telephone interview)	Measure: At each interview wave, respondents were asked to report whether their health was "excellent, very good, good, fair, or poor." Poor SRH if respondent self-rated poor or fair health (vs excellent or very good) in at least two consecutive interviews Specification: Binary (survival)

Nguyen et al., 2018	Twitter-derived sentiment toward racial and ethnic minoritized persons	Twitter API	n=1,249,653 tweets containing at least one race-related term	Sample: random 1% of geotagged Tweets from Twitter's API (March 2015–April 2016), subset Tweets referencing racial or ethnic groups/slurs using one or more of 398 race-related keywords Sentiment analysis: identified Tweets referencing black, Hispanic, Asian, White, and Middle Eastern groups and used machine learning algorithm with hand-coded training data to classify sentiment of Tweets: 1=positive, 0=negative/neutral Aggregate measure: state-level % of Tweets that was positive (racial minoritized persons overall, and broken out by racial group) Specification: tertiles (ref=T3)	LBW, VLBW, and PTB among birthing persons of various racial/ethnic groups	NCHS birth records	Administrative (birth) records	Measure: LBW: ≤ 2499g VLBW: ≤ 1499g PTB: gestational age < 37 weeks based on the obstetric estimate of gestation at delivery (OE). Specification: Binary for each outcome
Huang et al., 2020	Twitter-characterized sentiment toward racial and ethnic minoritized persons	Twitter API	n=30,977,747 tweets containing at least one race-related term	Sample: random 1% of geotagged/place-labeled Tweets from Twitter's API (2015-2018), subset Tweets referencing racial or ethnic groups/slurs using one or more of 518 race-related terms Sentiment analysis: used machine learning algorithm with hand-coded training data to classify sentiment of Tweets: negative (1=negative, 0=positive/neutral) and positive (1=positive, 0=negative/neutral) Aggregate measure: state-level % of Tweets that was negative and % that was positive Specification: Tertiles (ref=T1 for both)	CVD outcomes (e.g., hypertension, stroke) among various racial/ethnic groups	BRFSS	Self-report (telephone interview)	Question: Has a doctor, nurse or other health professional ever told you that you had … …hypertension, diabetes, obesity, stroke, myocardial infarction (MI), coronary heart disease (CHD)? Measure: Each outcome coded as binary (0=no, 1=yes). Any CVD if they answered "yes" to one or more. BMI: ≥30 kg/m2 was defined as obesity. Specification: Binary for each outcome
Nguyen et al., 2020	State-Level Racial Attitudes Assessed From Twitter Data	Twitter API	n=26,027,740 tweets from 2,498,717 Twitter users containing at least one race-related term	Sample: random 1% of geotagged/place-labeled Tweets from Twitter's API (June 2015-Dec 2017), subset Tweets referencing racial or ethnic groups/slurs using one or more of 518 race-related terms Sentiment analysis: identified Tweets referencing black, Hispanic, Asian, White, and Middle Eastern groups and used machine learning algorithm with hand-coded training data to classify sentiment of Tweets: negative (1=negative, 0=positive/neutral) and positive (1=positive, 0=negative/neutral) Aggregate measure: state-level % of Tweets that was negative and % that was positive	LBW and PTB among birthing persons of various racial/ethnic groups	NCHS birth records	Administrative data	Measure: LBW: ≤ 2499 g. PTB: gestational age < 37 weeks based on the obstetric estimate of gestation at delivery (OE) Specification: Binary

				<u>Specification</u>: Tertiles (ref=T1 for both)				
Hswen et al., 2020	Negative sentiment towards Mexicans and Hispanics during the 2016 presidential election	Twitter licensing agreement	n=2,809,641 tweets from 943,766 users containing terms Mexican(s) and/or Hispanic(s) (1,594,845 retweets)	<u>Sample:</u> full stream of tweets from Twitter over a 20-week period: 10 weeks before and 10 weeks after the 2016 United States presidential election <u>Sentiment analysis:</u> identified Tweets referencing Mexican(s) or Hispanic(s) (with and without #) and used VADER method to assign Tweets a continuous sentiment score ranging from -1 (most negative) to 1 (most positive), also collapsed into negative (< -0.5), positive (> +0.5), or neutral (-0.5 to +0.5). <u>Aggregate measure:</u> population-level weekly averages (whole US) <u>Specification</u>: weekly mean score and % negative, positive, and neutral	Daily negative mental wellbeing (worry)	Gallup-Sharecare Well-Being Index	Self-report (telephone interview)	<u>Measure</u>: Emotional well-being index measures Americans' daily experiences, and respondents categorize their responses as thriving, struggling, or suffering in the areas that measure wellbeing <u>Specification</u>: population-level weekly average % worry

Table B3. *Estimation and results*

Source	**Confounders**				**Estimation**		**Mediators/moderators evaluated**		
First author, year	**How identified**	**Area-level**	**Individual-level**	**How controlled**	**Statistical model**	**Adjusted MOA (95% CI or SE)**	**Mediators/ moderators**	**How assessed**	**Mediation/moderation findings**
Kennedy et al., 1995	Cited literature & defined as potential confounders: "some evidence suggests that low income and poverty are linked to depletion in social capital. Since income levels and poverty are also potential predictors of mortality, we evaluated these variables as potential confounders in the relationship between collective disrespect and mortality."	Median income, % in poverty	Accounted for age in the creation of rates	Multivariable regression	OLS regression	<u>Black mortality:</u> *No ability:* Beta=336.5, SE=93.4, p=0.0009 *No willpower*: Beta=256.1, SE=83.6, p=0.004 *Discrimination*: Beta=-290.1, SE=99.0, p=0.006 *Lack of educational opportunity*: Beta=-246.9, SE=83.9, p=0.006 <u>White mortality:</u> *No ability*: Beta=182.4, SE=71.9, p=0.01 *No willpower*: Beta=148.5, SE=62.5, p=0.02 *Discrimination*: Beta=-147.1, SE=75.0, p=0.06 *Lack of educational opportunity*: Beta=-173.8, SE=60.3, p=0.007 *Betas for one-unit change in collective disrespect*	Race (Black or White)	Examined race-specific mortality rates	Collective disrespect was associated with Black and White mortality rates but <u>results were stronger for Black mortality (>10% difference)</u>
Lee et al., 2015	Data-driven: All PSU-level covariates were chosen because they were significantly correlated with racial prejudice in bivariate models and there- fore could be potential confounders of the relationship	average number of people living below the federal poverty line (adjusted for family size and survey year), median income, average years of educational attainment, %	Race (White, Black), gender+, age at the time of the interview, marital status, household income, educational attainment	Multivariable regression	3-level HLM survival model	<u>Community-level racial prejudice, adjusting for individual-level prejudice and confounders:</u> OR=1.24; 95% CI=1.04, 1.49 *OR for 1SD change in community-level racial prejudice*	Moderators: Individual race (Black or White), Individual-level prejudice Mediator: community-level social capital	Multiplicative interaction terms in regression model: 1. race*individual-level prejudice; race*community-level prejudice 2. individual-level prejudice	<u>Race did not moderate</u> the association between community-level prejudice on mortality (race*prejudice interactions ns) or mediation through community-level social capital (described below) There was a significant interaction between <u>Individual*community-level prejudice</u>: OR = 0.74; 95% CI = 0.58,

	between racial prejudice and mortality Formation of candidate confounder list not specified	Black, located in the South, dissimilarity index, political affiliation index						*community-level prejudice Mediation: change-in-estimate approach	0.95, indicating that individuals low in racial prejudice but living in higher-prejudice communities had the highest level of mortality risk. <u>Mediation</u>: Social capital was inversely related to community-level prejudice (r =–0.41; P <.01), indicating that communities with higher levels of prejudice had lower levels of social capital. When social capital was controlled in the fully adjusted model, PSU-level racial prejudice was no longer significantly associated with mortality.
Morey et al., 2018	Prior research: We included variables that prior research suggested may be potential confounders of the association between anti-immigrant prejudice and mortality.	% foreign-born, mean years of education, mean family income, % who identify as politically conservative, survey year for anti-immigrant score	Gender,[+] age, marital status , years of education, unemployment, family income, self-rated health at baseline	Multivariable regression	Cox proportional hazards models with clustered SEs	<u>Community-level anti-immigrant prejudice and mortality main effects</u>: HR=1.05 95% CI=0.93, 1.19 (ns) *HR for 1-unit change in anti-immigrant prejudice score*	Nativity status, Race (Black, White, Other – *sensitivity analysis restricted "other race" to Asian and Hispanic)*	Multiplicative interaction terms in regression model: race*community-prejudice; nativity*community-prejudice; race*nativity*community-prejudice; Also stratified results by race (Table 3)	<u>Race*nativity moderated</u> nativity*prejudice: ns race*prejudice: ns race*nativity*prejudice: sig (F-test=4.04, p=0.018) – interpretation: *the association between anti-immigrant prejudice and mortality for US-born respondents was significantly different compared to foreign-born respondents* <u>Stratified findings by race and nativity:</u> <u>US-born "other race":</u> The mortality hazard ratio for US-born respondents living in high-prejudice communities (HR=2.63 [95% CI: 0.53, 13.12]) was 171% higher than US-born respondents living in low-prejudice

									communities (HR=1.54 [95% CI: 0.75, 3.18]). Foreign-born "other race": The mortality hazard ratio for foreign-born respondents living in the high-prejudice communities (HR=0.15 [95% CI: 0.02, 1.20]) was 287% lower than foreign-born respondents living in low-prejudice communities (HR=0.43 [95% CI: 0.17, 1.09]). *HR comparing mortality in high (1SD above the mean) vs low (1SD below the mean) prejudice communities* Results restricted to Asian and Hispanic "other race" respondents showed similar patterns but were less precise due to small number of respondents.
Leitner et al., 2016a	Not stated with two exceptions: (1) Geomobility: "Importantly, a relationship between Blacks' racial bias and ingroup health could be driven by social selection forces." (2) Age bias: "To examine whether any effects were specific to racial bias, or generalized to bias on	Study 1: total population, Black-to-White ratio, dissimilarity index of segregation, Black geographic mobility, housing density, urbanicity (number of housing units per square mile), implicit and explicit age bias, and average of	Study 1: accounted for sex, age, and race in creation of rates Study 2: accounted for age and race in creation of rates	Multivariable regression	GEE with robust standard errors and simple slopes	White explicit bias and circulatory disease death rates: Black death rates (positive, stronger): b=43.20, SE=12.10, p=0.0004, White death rates (positive, weaker): b=13.90, SE=4.97, p=0.005 Implicit bias ns (simple slopes estimates not shown) *b for 1-point increase in racial bias of White people*	Effect modification: Race (Black or White) Mediation: Black-White disparities in health behaviors (smoking, drinking, and exercise)	Multiplicative interaction term in regression model: Race*White implicit bias Race*White explicit bias (also 3-way interaction with race*sex, but results were ns) Sig interaction effects	Study 1 – NA (no main effects on circulatory disease diagnosis) Study 2 – Race moderated association between explicit racial bias and healthcare access (sig race*implicit bias interaction). Simple slopes: Whites' explicit racial bias was associated with White and Black circulatory disease death rates, but stronger association with Black rate (race*implicit ns)

	nonracial dimensions”	Black and White: high school graduation rates, MHI past 12 months, unemployment, % in poverty *each covariate interacted with individual race Study 2: same as study 1 + neoplasm (cancer) death rate						explored via simple slopes analysis Mediation: Change-in-estimates approach	Mediation: Black-White disparities in health behaviors did not mediate the relationship between explicit/implicit bias and death rate disparity.
Leitner et al., 2016b	“We adopted an analytic approach that could test whether Blacks’ bias remained a predictor of Blacks’ death rate when we controlled for a large set of socio-demographic characteristics and whites’ biases in the same county.” Further explicated rationale for sex ratio (previous research), income inequality (previous research), and geomobility	Total population, Black population, Black/White MHI, Black/White high school graduation rate, Black/White poverty rate, Black/White unemployment rate, dissimilarity index of segregation, housing density, Black geographic mobility, income inequality, Black/White male-to-female ratio,	NA	Multivariable regression	Analysis 1: GEE with robust SEs and simple slopes	Black ingroup bias and Black death rate: *Explicit*: b=0.005, SE=6.20, Beta <0.001, p=0.99 *Implicit:* b=157.24, SE=34.04, Beta =0.49, p<0.0001 White ingroup bias and White death rate: *Explicit:* b=19.04, SE=4.98, p=0.0001 *Implicit*: b=23.81, SE=28.10, p=0.40 Note: estimates derived from simple slope analysis with race*ingroup bias interactions *b for 1-point increase in ingroup bias*	Race (Black or White)	Multiplicative interaction term in regression model: Race*ingroup implicit bias Race*ingroup explicit bias (also looked at higher order interactions with ingroup implicit*explicit *race, but results were ns) Sig interaction effects explored via simple slopes analysis	Race moderated the association between implicit and explicit racial bias and ingroup death rates: Implicit ingroup bias was associated with Black but not White death rates Explicit ingroup bias was associated with White but not Black death rates

	(conjecture, content knowledge)	ingroup racial bias of other group (e.g., models for Black bias and health controlled for Whites' implicit and explicit ingroup bias)							
Orchard & Price, 2017	"We include additional covariates to reduce the possibility of county-level prejudice being correlated with other individual and county characteristics."	Total population, unemployment rate, % college graduates, % Black*, Black poverty rate*, sexual orientation IAT, gender-career IAT * interacted with birthing person's race	Maternal age, marital status, education, and 17 different pregnancy risk factors (e.g., high blood pressure, previous preterm birth, etc.); child gender+ and birth order of the child.	Multivariable regression	Weighted least squares regression with clustered SEs and state- and year- fixed effects	*Implicit:* LBW: The black-White gap in low birth weight is 14% larger in counties with high vs low implicit racial prejudice. PTB: The black-White gap in low birth weight is 29% larger in counties with high vs low implicit racial prejudice. *Explicit*: LBW: The black-White gap in low birth weight is 22% larger in counties with high vs low explicit racial prejudice. PTB: The black-White gap in low birth weight is 36% larger in counties with high vs low explicit racial prejudice. Notes: (1) When implicit and explicit bias were modeled together, only explicit remained significant predictor of B-W birth outcome gaps. (2) Explicit prejudice in county of birth more strongly associated with B-W birth outcome gaps than county of residence (results similar for implicit). (3) Results unique to racial bias (no findings for gender-career or sexual orientation bias)	Birthing person's race (Black or White); County of residence vs. county of birth	Multiplicative interaction term in regression model: county prejudice*birthing person's race. Used models to estimate race-specific effects, and plotted stratified results. Also stratified results on bias in county of residence vs. birth county	Race moderated the association between community-level prejudice and birth outcomes (sig interaction). Findings showed stronger associations among Black birthing persons and no (or even protective) associations among White birthing persons. Prejudice in community of birth was more strongly associated with birth outcomes than prejudice in community of residence.

Hehman et al., 2017	Inflated model used to "develop initial predictive models of lethal force" (supplemental material include more parsimonious models determined using data-driven approach: forward and backward stepwise regression)	Black/White MHI, Black/White % with HS or equivalent degree, isolation index of segregation, violent crime, unemployment, population density, and total (race disaggregated) lethal force	NA	Multivariable regression	Linear regression	<u>Model with race-IAT</u> White *implicit*: b=4.13, SE=1.90, p=0.031 White *explicit*: b=-0.52, SE=0.29, p=0.079 Black *implicit*: b=-1.13, SE=0.84, p=0.182 Black *explicit*: b=0.12, SE=0.14, p=0.40 <u>Model with race-IAT and weapons-IAT</u> Implicit/explicit racial bias of White and Black respondents – ns White *implicit threat stereotypes*: b=5.50, SE=1.63, p=0.001 *b for 1-point increase in race- and weapons-IAT of White people*	NA	NA	NA, but note they did calculate a disproportionate lethal force measure for White people and found they were not being killed disproportionately. Therefore, estimation was just for the association between regional racial bias and disproportionate killing of Black people.
Chae et al., 2015	Adjusted for "relevant area-level covariates"	% in urbanized area (>50,000 population), % Black, % Blacks with up to a high school education, % Black households in poverty, White mortality rate	Accounted for age group, sex, year of death, census region in creation of rates	Multivariable regression	Negative binomial regression model with Huber-White clustered SEs	<u>All cause</u>: MRR=1.04, 95% CI = 1.02, 1.06 <u>Heart disease</u>: MRR=1.04, 95% CI=1.02, 1.07 <u>Cancer</u>: MRR=1.03, 95% CI=1.00, 1.05 <u>Stroke</u>: MRR=1.03, 95% CI=1.00, 1.07 <u>Diabetes</u>: MRR=0.95, 95% CI=0.88, 1.019 *MRR for 1SD increase Google searches for N-word*	NA	NA	NA
Chae et al., 2018	"conceptual relevance" + data-driven (changes-in-estimates) discussed	Census region, % Black, % in urbanized area (>50,000 population), % Black w/ <HS degree or equivalent,	Maternal age	Multivariable regression	Log-binomial regression model fit with GEE	<u>PTB:</u> PR=1.05, 95% CI=1.02, 1.09, <u>LBW:</u> PR=1.05, 95% CI=1.02, 1.07 *PR for 1SD increase in Google searches for N-word*	NA	NA	NA

		% Black in poverty							
McKetta et al., 2018	Adjusted for "relevant confounders"	Median income, % Black population (sensitivity), Google searches for N-word in SRH->movement model	At baseline: SRH, age, education level	Multivariable regression	Incident SRH: Cox PH Movement: Logistic regression	Incident SRH among White respondents: Q2: HR=1.19, 95% CI=1.07, 1.32 Q3: HR=1.13, 95% CI=1.04, 1.22 Q4: HR=1.33, 95% CI=1.20, 1.47 Incident SRH among Black respondents: Q2: HR=1.43, 95% CI=1.12, 1.82 Q3: HR=1.31, 95% CI=1.05, 1.63 Q4: HR=1.20, 95% CI=0.95, 1.50 *Ref = Q1 racial animus (Google searches for N-word)*	Race (Black or White)	Multiplicative interaction term in regression model: race*state-level racial animus	Race did not moderate the association between state-level racial animus and poor SRH (interaction term ns)
Nguyen et al., 2018	Individual: "to adjust for potential confounding of the relationship between neighborhood environments and birth outcomes." State: "to account for between-state differences in compositional characteristics."	MHI, % NH White	Maternal age, marital status, race, Hispanic ethnicity, education, BMI, smoking status during pregnancy, first birth indicator, prenatal care in the 1st trimester indicator	Multivariable regression	Log Poisson regression models with robust SEs	T1 vs T3 positive sentiment toward racial/ethnic minoritized persons LBW: PR=1.06, 95% CI=1.04, 1.07 VLBW: PR=1.09, 95% CI=1.06, 1.12 PTB: PR=1.10, 95% CI=1.10, 1.11 Note: sentiment towards specific racial/ethnic groups showed a similar pattern of results	Birthing person's race/ethnicity (White vs Hispanic or nonWhite or foreign-born)	Stratified subgroup analyses (did not test for statistical interaction)	Race/ethnicity did not moderate association between Twitter sentiment and birth outcomes: Results from subgroup analyses restricted to racial/ethnic minorized birthing persons did not differ substantially from those seen for the full population of birthing persons (differences in PRs <10%).
Huang et al., 2020	Not stated	% NH White, % NH Black, % Hispanic, MHI	Age, sex, education, race/ethnicity, and marital status	Multivariable regression	Poisson regression	T3 vs T1 negative sentiment toward racial/ethnic minoritized persons Hypertension: PR=1.11, 95% CI=1.08, 1.14 Diabetes: PR=1.15, 95% CI=1.08, 1.22 Obesity: PR=1.14, 95% CI=1.10, 1.18 Stroke: PR=1.30, 95% CI=1.16, 1.46	gender+ and race/ethnicity	Assessed statistical interactions: Sentiment*sex Sentiment*race/ethnicity	Race and sex did moderate, but findings depended on the outcome: + In general, effects were stronger for women (except on diabetes and obesity) + Negative sentiment and hypertension, MI, and any CVD = stronger for non-Hispanic Whites and non-Hispanic blacks than other race/ethnicity

						MI: PR=1.14, 95% CI=1.03, 1.25 CHD: PR=1.09, 95% CI=1.00, 1.19 Any CVD: PR=1.16, 95% C=1.09, 1.24 T3 vs T1 positive sentiment toward racial/ethnic minoritized persons Hypertension: PR=0.97, 95% CI 0.94, 1.00 Diabetes: PR=0.94, 95% CI 0.90, 0.99 Obesity: PR=0.97, 95% CI 0.94, 1.00 Stroke: PR=0.89, 95% CI 0.80, 0.98 MI: PR=0.91, 95% CI 0.83, 0.98 CHD: PR=0.94, 95% CI 0.86, 1.02 Any CVD: PR=0.90, 95% CI 0.86, 0.95			groups + Negative sentiment and diabetes, obesity, and stroke = stronger in Hispanics than any other racial/ethnic groups + Positive sentiment and hypertension, diabetes, and obesity = effects more protective in non-Hispanic Black than non-Hispanic Whites
Nguyen et al., 2020	Individual: "We adjusted for potential confounders of the association between racial sentiment and birth outcomes." State: "to account for state-level compositional differences in demographic and economic characteristics."	% NH Black, % Hispanic, population density, Southern state indicator, economic disadvantage composite. (% unemployed; % some college education, % high school diploma, % children in poverty, % single parent household, MHI)	Maternal age, marital status, race, Hispanic ethnicity, education, BMI, smoking status during pregnancy, first birth indicator, prenatal care in the 1st trimester indicator, birth year	Multivariable regression	log binomial regression models with clustered SEs	LBW: T2: IR=1.08, 95% CI=1.03-1.13 T3: IR=1.08, 95% CI=1.04-1.13 PTB: T2: IR=1.09, 95% CI=1.04-1.13 T3: IR=1.08, 95% CI=1.00-1.14 *Ref=T1 negative sentiment toward racial/ethnic minoritized persons*	Birthing person's race (Black NH, White NH, Asian NH, Hispanic, and all minoritized persons)	Stratified subgroup analyses (did not test for statistical interaction)	Race did not moderate: State-level sentiment toward all minoritized people was associated with adverse birth outcomes among all birthing persons (differences in IRs <10%). Negative sentiment toward racial/ethnic minoritized persons (T3 vs T1 (ref)): *Among all racial/ethnic minoritized birthing persons:* LBW: IR=1.13 (1.06-1.21) PTB: IR=1.10 (1.05-1.16) *Among WhiteWhite birthing persons:* LBW: IR=1.08 (1.03-1.14)

									PTB: IR=1.08 (1.00-1.17) Also examined race-concordant associations (e.g., sentiment toward Hispanics and outcomes among Hispanic birthing persons, so not effect modification per-se, but results showed subgroup effects for Black and Middle Eastern birthing persons) Also, for Black birthing persons (vs full sample) the associations between negative Twitter sentiment toward Black people and birth outcomes became stronger over time (2015<2016<2017)
Hswen et al., 2020	NA	NA	NA	NA	Time series lag (autoregressive distributed) regression model	LR lag = 0.31; p = 0.022 *Interpretation: Negative tweets mentioning Mexicans and Hispanics predicted daily worry with significant lag time of one week.*	NA	NA	NA

Abbreviations:

Data sources:

GSS: General Social Survey

NDI: National Death Index

IAT: Implicit Association Test

API: Application Program Interface

BRFSS: Behavioral Risk Factor Surveillance Survey

PSID: Panel Study of Income Dynamics

NCHS: National Center for Health Statistics

Geographic scales:

DMA: designated market area (media markets receiving similar media and news programming)

CBSA: Core-based statistical area (similar to metropolitan areas)

PSU: Primary sampling units (metropolitan statistical areas and nonmetropolitan counties)

Estimation:

MV: multivariable

GEE: generalized estimating equation

OLS: ordinary least squares

HLM: Hierarchical linear model

PH: proportional hazard

OR: odds ratio

PR: prevalence ratio

IR: incidence ratio

MRR: mortality rate ratio

ns: not statistically significant ($p \geq 0.05$)

Study measures:

MHI: median household income

HS: high school

NH: non-Hispanic

BMI: body mass index

PTB: preterm birth

LBW: low birthweight

VLBW: very low birthweight

CVD: cardiovascular disease

CHD: coronary heart disease

MI: myocardial infarction

Other:

E: exposure

O: outcome

C: covariates

Data source information:

GSS: The General Social Survey (GSS) is a nationally-representative sample of non-institutionalized English-speaking adults aged 18+ living in the United States conducted on a new population sample at each wave (Davis & Schwartzman, 1973; Kennedy et al., 1997; Lee et al., 2015; Morey et al., 2018).

BRFSS: The Behavioral Risk Factor Surveillance System (BRFSS) is a telephone-based survey (random-digit dialing of landlines and cellphones) is a telephone-based survey that focuses on chronic health conditions and health behaviors of adults across 50 states of USA and District of Columbia ((CDC); Huang et al., 2020).

PSID: The Panel Study on Income Dynamics (PSID) is a nationally representative, longitudinal study of households in the U.S. with interviews collected biannually by phone (Center; McKetta et al., 2017).

Project Implicit: Project Implicit (PI) is a Harvard-based nonprofit research project which provides a free, online tool for assessing implicit and explicit biases toward various social groups (e.g., Black vs. White persons, gay vs. straight persons) (Nosek et al., 2010). PI measures explicit biases via self-report and implicit biases via the "Implicit Association Test" (IAT).

Implicit Association Test with D-scoring algorithm: The "Implicit Association Test" (IAT) is a speeded dual-categorization task which measures the speed of keyboard associations between images of Black vs. White faces and positive (e.g., wonderful) vs. negative (e.g., disgusting) words. Faster reaction time matching positive words with White and negative words with Black faces indicates cognitive dissonance between Black people and positive emotions, which is interpreted as a pro-White implicit bias and/or anti-black Bias (Hehman et al., 2018; Leitner et al., 2016a, 2016b). The IAT is scored using the D-score measure, which ranges from -2 to +2 (Greenwald et al., 2003).

Explicit temperature explicit measure: Two feeling thermometer items separately ask how warm or cold participants feel toward both African Americans and European Americans (0 = very cold, 10 = very warm). Responses to the Black feeling thermometer are subtracted from responses to the White feeling thermometer, creating a score that ranges from -10 to +10 with higher values representing warmer feelings toward White people compared to Black people, interpreted as a pro-White/anti-Black explicit bias (Hehman et al., 2018; Leitner et al., 2016a, 2016b).

Explicit preference measure: respondents describe how they feel toward European and African Americans using a scale that ranges from "I strongly prefer African Americans to European Americans", to "I strongly prefer European Americans to African Americans." Responses are on a 5-point Likert scale until 2006 and a 7-point Likert scale after 2006.(Orchard & Price, 2017)

Notes:

Data on Google Searches for the N-word from 2004-2007 were extracted by Seth Stephens-Davidowitz (2014) using an older version of the Google Trends platform (the algorithm has since changed).

Twitter data were all geolocated with either geotag (latitude and longitude) only (Nguyen et al., 2018) or geotag and user-provided "place" information (Huang et al., 2020; Nguyen et al., 2020).

+ conflated sex and gender (i.e., stated they measured gender but variables were male/female (i.e., biologic sex))

Appendix C. *Overview of data sources used to measure area-level racial prejudice*

Data source	Used in studies	Years available	Geographies available	Data access	Number aggregated	Indicator of racial prejudice
General Social Survey	Kennedy et al., 1997; Lee et al., 2015; Morey et al., 2018	1972-2018, collected every 3 years (racial attitudes questions asked beginning in 1993)	State, PSU, county, census tract	Restricted – must apply for data	Ranges from about 2,500 to 13,355 across studies	<u>Anti-Black racial prejudice:</u> Composite score based on questions: "On average, blacks have worse jobs, income, and housing. Do you think the differences are due to (a) discrimination, (b) less in-born ability to learn, (c) lack of chance for education that it takes to rise out of poverty, (d) less motivation or willpower to pull themselves out of poverty?" (Kennedy et al., 1997), "Do blacks tend to be unintelligent or tend to be intelligent?", and "Do blacks tend to be hard working or lazy?" (Lee et al., 2015) <u>Anti-immigrant prejudice:</u> Composite score based on questions: "Do you think the number of immigrants to America nowadays should be increased a lot, increased a little, remain the same as it is, reduced a little, or reduced a lot?" and agree or disagree with the following statements: (1) "America should take stronger measures to exclude illegal immigrants," (2) "Immigrants take jobs away from people who were born in America," (3) "Immigrants increase crime rates," and (4) "Immigrants are generally good for America's economy" (Morey et al., 2018).
Project Implicit	Leitner et al., 2016a; Leitner et al., 2016a; Orchard & Price,	2002 - present, collected continuously	State, CBSA, county	Publicly available	Ranges from about 250,000 to 1.8 million across studies	<u>Pro-White/anti-Black racial prejudice:</u> Implicit – assessed using the Implicit Association Test

	2017; Hehman et al., 2018					Explicit – assessed via self-report (temperature measure or preference measure) All measures scored so negative values imply pro-Black/anti-White bias, positive imply pro-White/anti-Black bias, and 0 implies a neutral score.
Google Trends	Chae et al., 2015; Chae et al., 2018; McKetta et al., 2018	2004 - present, collected continuously	State, DMA, some cities	Publicly available	NA	Relative popularity of Google searches containing the “n-word” (ending in “-er(s)” but not “-a(s)”) (Chae et al., 2015; Chae et al., 2018). Scored on a scale from 1 to 100 where the region with the highest search volume over the study period is assigned a score of 100 and all other regions are given a relative score. “Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term” (Google, 2020)
Twitter	Nguyen et al., 2018; Huang et al., 2020; Nguyen et al., 2020; Hswen et al., 2020	2006 – present, with option for retrospective or prospective collection	Latitude + longitude available for 3-4% of public tweets, state information discernable for ~ 99% of tweets	Publicly available	1 million – 30 million	Proportion of public Tweets with latitude and longitude or other “place” information (e.g., city, state) referencing a particular racial/ethnic group that are positive, negative, or neutral. Sentiment is determined based on a combination of hand-coding, natural language processing, and machine learning.

Geographic scales:

DMA: Designated Market Area (media markets receiving similar media and news programming)

CBSA: Core-based statistical area (similar to metropolitan areas)

PSU: Primary sampling units (metropolitan statistical areas and nonmetropolitan counties)

Appendix D. Strengths and limitations of data sources used to measure area-level racial prejudice

Data source	Primary strengths	Primary limitations
General Social Survey	Nationally representative Racial bias questions have been asked since 1993, offering greater historical context compared to the other measures* Specificity in measurement: questions ask directly about racial attitudes* Information on demographics of respondents (e.g., race, age, political identification, etc) is available*	Not all questions are asked to all participants or on all survey years Social desirability – because racial attitudes are self-reported, the GSS is subject to self-censorship or social desirability bias Must apply for data*
Project Implicit	Over 3-million tests have been taken since 2002 Publicly available and free Multiple validated tests available (e.g., racial bias, age bias, gender bias, etc.) Can disentangle implicit vs explicit bias Circumvent social desirability/self-censorship: IAT measures *implicit* bias through keyboard association test which does not rely on self-report Information on demographics of test-takers (e.g., race, age, political identification, etc.) is available	Project Implicit respondents are self-selected and therefore racial bias cannot be generalized to any broader population (note: some studies apply post-stratification weights on age/sex but non-representativeness on other dimensions may persist) Repeat test-takers may regress toward the mean*

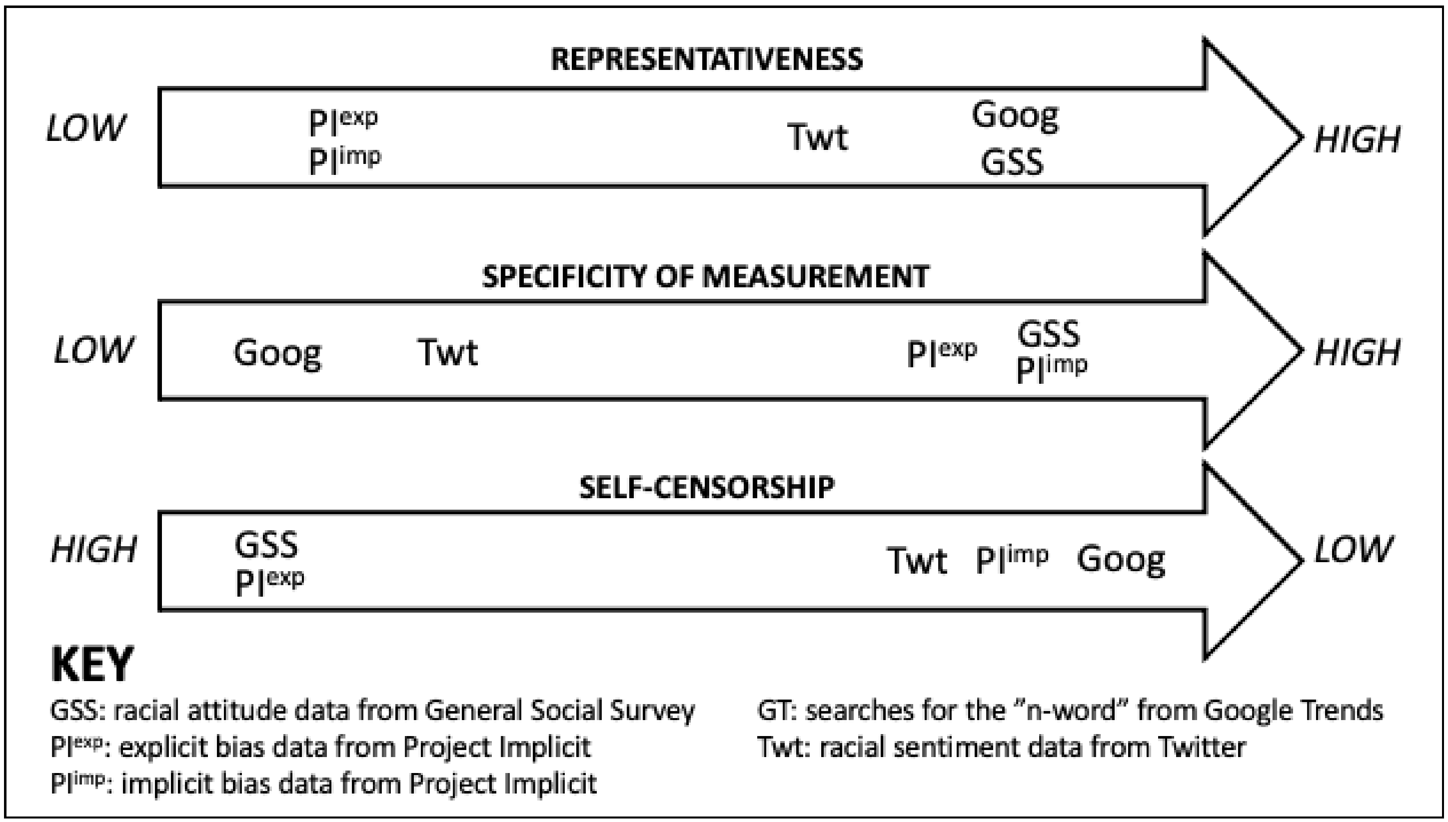
	Has shown high convergent validity with other measures of area-level bias*	
Google Trends	Widely and regularly used by many people around the world Circumvent social desirability/self-censorship: does not rely on self-report and search data captures *private curiosities* Allows for real-time analysis of social attitudes* Has been used for disease surveillance and prediction Has shown high convergent validity with other measures of area-level bias	Context of the search is unknown Internet queries for the "N-Word" may not be motivated by racism Demographics of person conducting the search are unknown* Not possible to discern multiple searches from the same user*
Twitter	Widely and regularly used by many people around the world Millions of tweets are sent daily and over 90% of Twitter users make their profile and communication public Circumvent *some* social desirability/self-censorship: does not rely on self-report and sense of anonymity may embolden users to express views they would not display during in-person interactions Allows for real-time analysis of social attitudes Sentiment analysis allows researcher to characterize Tweets as positive, negative, or neutral	Geolocation data only available for small proportion of tweets where user either a) enables latitude + longitude or b) shares location of Tweets – may lead to systematic bias Potential for residual self-censorship*:* Twitter only reflects what people were willing to express publicly Sentiment analysis unable to identify and process sarcasm or humor in a tweet Demographics of person writing the Tweets are unknown

	Has been used to characterize sentiment around a number of health topics and health outcomes	

Note: Information in this table is extracted from the 14 papers included in the systematic review. Any information that comes from content area knowledge or outside literature is indicated with an *

Figure D.

Measurement Trade-Offs Between Area-Level Racial Prejudice Data Sources



REFERENCES

* indicates studies included in the systematic review

(CDC), C. f. D. C. a. P. *Behavioral Risk Factor Surveillance System Survey*.

Center, S. R. *Panel Study of Income Dynamics*.

* Chae, D. H., Clouston, S., Hatzenbuehler, M. L., Kramer, M. R., Cooper, H. L., Wilson, S. M., Stephens-Davidowitz, S. I., Gold, R. S., & Link, B. G. (2015). Association between an internet-based measure of area racism and Black mortality. *PLoS ONE, 10*(4), e0122963.

* Chae, D. H., Clouston, S., Martz, C. D., Hatzenbuehler, M. L., Cooper, H. L., Turpin, R., Stephens-Davidowitz, S., & Kramer, M. R. (2018). Area racism and birth outcomes among Blacks in the United States. *Social Science & Medicine, 199*, 49-55.

Davis, J. A., & Schwartzman, K. (1973). *General Social Survey: March 1975* (Vol. 4). Inter-University Consortium for Political Research.

Google. (2020). *Google Trends*. https://trends.google.com/trends/?geo=US

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of personality and social psychology, 85*(2), 197.

* Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science*, 1948550617711229.

* Hswen, Y. (2020). Online negative sentiment towards Mexicans and Hispanics and impact on mental well-being: A time-series analysis of social media data during the 2016 United States presidential election. *Heliyon, 6*(9). https://doi.org/10.1016/j.heliyon.2020.e04910

* Huang, D., Huang, Y., Adams, N., Nguyen, T. T., & Nguyen, Q. C. (2020). Twitter-Characterized Sentiment Towards Racial/Ethnic Minorities and Cardiovascular Disease (CVD) Outcomes. *Journal of racial and ethnic health disparities*, 1-13.

Innovation, V. H. (2016). *Covidence systematic review software.* In www.covidence.org

* Kennedy, B. P., Kawachi, I., Lochner, K., Jones, C., & Prothrow-Stith, D. (1997). (Dis) respect and black mortality. *Ethnicity & disease, 7*(3), 207-214.

* Lee, Y., Muennig, P., Kawachi, I., & Hatzenbuehler, M. L. (2015). Effects of racial prejudice on the health of communities: a multilevel survival analysis. *American Journal of Public Health, 105*(11), 2349-2355.

* Leitner, J. B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016a). Blacks' death rate due to circulatory diseases is positively related to whites' explicit racial bias: A nationwide investigation using project implicit. *Psychological Science, 27*(10), 1299-1311.

* Leitner, J. B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016b). Racial bias is associated with ingroup death rate for Blacks and Whites: Insights from Project Implicit. *Social Science & Medicine, 170*, 220-227.

* McKetta, S., Hatzenbuehler, M. L., Pratt, C., Bates, L., Link, B. G., & Keyes, K. M. (2017). Does social selection explain the association between state-level racial animus and racial disparities in self-rated health in the United States? *Annals of Epidemiology, 27*(8), 485-492. e486.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med, 6*(7), e1000097.

* Morey, B. N., Gee, G. C., Muennig, P., & Hatzenbuehler, M. L. (2018). Community-level prejudice and mortality among immigrant groups. *Social Science & Medicine, 199*, 56-66.

* Nguyen, T. T., Adams, N., Huang, D., Glymour, M. M., Allen, A. M., & Nguyen, Q. C. (2020). The Association Between State-Level Racial Attitudes Assessed From Twitter Data and Adverse Birth Outcomes: Observational Study. *JMIR Public Health and Surveillance, 6*(3), e17103.

* Nguyen, T. T., Meng, H.-W., Sandeep, S., McCullough, M., Yu, W., Lau, Y., Huang, D., & Nguyen, Q. C. (2018). Twitter-derived measures of sentiment towards minorities (2015–2016) and associations with low birth weight and preterm birth in the United States. *Computers in Human Behavior*.

Nosek, B., Banaji, M., & Greenwald, A. (2010). Project implicit. *Project Implicit*.

* Orchard, J., & Price, J. (2017). County-level racial prejudice and the black-white gap in infant health outcomes. *Social Science & Medicine, 181*, 191-198.

Software, V. (2019). *MAXQDA 2020.* In maxqda.com

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics, 118*, 26-40.